**List Pertanyaan**

|  |  |
| --- | --- |
| **PERTANYAAN** | **JAWABAN DAN PENJELASAN** |
| Terminology of data | Data sesuatu yang belum mempunyai arti bagi penerimanya dan masih memerlukan pengolahan. Contoh : gambar, suara, huruf, angka dan lain lain |
| Statistics, sample, populasi | - Statistics, adalah ukuran yang berlaku bagi sampel  - Sample,adalah sebagian dari populasi yang digunakan sebagai sumber data penelitian.  - Populasi, adalah |
| Sampling method | Metode Teknik sampling adalah merupakan teknik pengambilan sampel. Contoh : sampling nonprobabilitas, sampling probabilitas |
| Data measurement (nominal, ordinal, interval, ratio) |  |
| Correlation (pearson vs. Spearman) |  |
| Regression, base line model, assumption |  |
| Deductive vs. Inductive reasoning (including example) |  |
| Big data terminology | Big Data adalah istilah yang menggambarkan volume data yang besar, baik data yang terstruktur maupun data yang tidak terstruktur. |
| CRISP-DM or other analytics process |  |

**Input:**

TelkomIndonesia.json

UseeTVcom.json

**Tugas:**

1. Eksplorasi ‘text’ dari kedua file json di atas
2. Carilah karakter pertama dari setiap kata dari text-text tersebut. Abaikan tanda baca.
3. Untuk setiap karakter pertama tersebut, hitunglah:
4. word\_count: total kata yang diawali karakter tersebut
5. hashtag\_count: total hashtag yang diawali karakter tersebut
6. mention\_count: total mention yang diawali karakter tersebut
7. top3\_words: tiga kata yang paling banyak muncul beserta frekuensinya
8. top3\_hashtag: tiga hashtag yang paling banyak muncul beserta frekuensinya
9. top3\_mention: tiga akun yang paling sering di-mention beserta frekuensinya
10. Bisa gunakan Hive, Spark, R, Python, Scala
11. Contoh output:

{

"first\_letter" : "a",

"word\_count" : 1000,

"hashtag\_count" : 100,

"mention\_count" : 50,

"top3\_words" : [{"word" : "abang", "count" : 150}, {"word" : "adek", "count" : 100}, {"word" : "adadeh", "count" : 50}],

"top3\_hashtags" : [{"hashtag" : "#abc", "count" : 15}, {"hashtag" : "#aku", "count" : 10}, {"hashtag" : "#alamak", "count" : 8}],

"top3\_mentions" : [{"mention" : "@amerika", "count" : 8}, {"mention" : "@argentina", "count" : 5}, {"mention" : "#australi", "count" : 2}]

}

**JAWAB**

**# Extracting and sampling**

**# Setter Default and ignored boolean / statement**

**null = None**

**false = False**

**true = True**

**# Base Array**

**datanyah = {**

**'base': [**

**{"created\_at":"Tue Nov 07 07:00:01 +0000 2017","id":9.27792741819601e+017,"id\_str":"927792741819600897","text":"Semua orang bisa membuat video yang berkualitas lho, yuk pelajari tips merekam video secara profesional menggunakan smartphone! #TelkomEdu","truncated":false,"entities":{"hashtags":[{"text":"TelkomEdu","indices":[128,138]}],"symbols":[],"user\_mentions":[],"urls":[]},"source":"<a href=\"http://sproutsocial.com\" rel=\"nofollow\">Sprout Social<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":80204488,"id\_str":"80204488"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":2,"favorite\_count":0,"favorited":false,"retweeted":false,"lang":"in"},**

**{"created\_at":"Tue Nov 07 06:00:11 +0000 2017","id":9.27777684238815e+017,"id\_str":"927777684238815232","text":"Petualangan seorang bocah laki-laki bernama Wally Trollman dan naga kesayangan, Norville. Wallykazam 15.05 WIB Nick… https://t.co/xI79lkpTux","truncated":true,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/xI79lkpTux","expanded\_url":"https://twitter.com/i/web/status/927777684238815232","display\_url":"twitter.com/i/web/status/9…","indices":[117,140]}]},"source":"<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">Hootsuite<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":0,"favorite\_count":0,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},**

**{"created\_at":"Tue Nov 07 04:00:10 +0000 2017","id":9.27747477796741e+017,"id\_str":"927747477796741120","text":"Aktifkan HOOQ dari UseeStore &amp; dapatkan bonus TVOD film &amp; serial TV Hollywood setiap bulannya di @HOOQ\_ID!… https://t.co/yQMjHMGAka","truncated":true,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[{"screen\_name":"HOOQ\_ID","name":"HOOQ Indonesia","id":3322010498,"id\_str":"3322010498","indices":[105,113]}],"urls":[{"url":"https://t.co/yQMjHMGAka","expanded\_url":"https://twitter.com/i/web/status/927747477796741120","display\_url":"twitter.com/i/web/status/9…","indices":[116,139]}]},"source":"<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">Hootsuite<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":0,"favorite\_count":0,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},**

**{"created\_at":"Mon Nov 06 12:00:24 +0000 2017","id":9.27505947039338e+017,"id\_str":"927505947039338497","text":"Saksikan rangkaian pertandingan Golf terkini secara gratis hingga 15 Nov 2017 di Golf Channel &amp; Golf+ Subscribe now… https://t.co/1LkYAhQDT8","truncated":true,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/1LkYAhQDT8","expanded\_url":"https://twitter.com/i/web/status/927505947039338497","display\_url":"twitter.com/i/web/status/9…","indices":[121,144]}]},"source":"<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">Hootsuite<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":0,"favorite\_count":0,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},**

**{"created\_at":"Mon Nov 06 08:00:16 +0000 2017","id":9.27445515125645e+017,"id\_str":"927445515125645312","text":"Yuk ikutan kuis #HOOQOriginals Premier Film Marlina Si Pembunuh dlm Empat Babak! Jwb pertanyaannya di… https://t.co/1Fo1P06HUA","truncated":true,"entities":{"hashtags":[{"text":"HOOQOriginals","indices":[16,30]}],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/1Fo1P06HUA","expanded\_url":"https://twitter.com/i/web/status/927445515125645312","display\_url":"twitter.com/i/web/status/9…","indices":[103,126]}]},"source":"<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">Hootsuite<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":0,"favorite\_count":0,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},**

**{"created\_at":"Mon Nov 06 04:29:49 +0000 2017","id":9.27392553103708e+017,"id\_str":"927392553103708160","text":"Dont Breathe - Suspense Thriller - HBO: https://t.co/tOyOdcJ27Y via @YouTube","truncated":false,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[{"screen\_name":"YouTube","name":"YouTube","id":10228272,"id\_str":"10228272","indices":[68,76]}],"urls":[{"url":"https://t.co/tOyOdcJ27Y","expanded\_url":"http://youtu.be/bHUnHmFVXMs?a","display\_url":"youtu.be/bHUnHmFVXMs?a","indices":[40,63]}]},"source":"<a href=\"http://www.google.com/\" rel=\"nofollow\">Google<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":1,"favorite\_count":1,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"en"},**

**{"created\_at":"Mon Nov 06 04:00:10 +0000 2017","id":9.27385091889861e+017,"id\_str":"927385091889860609","text":"Mau kado natal dari IndiHome? Ikuti promo undian IndiHome Paket Natal &amp; Tahun Baru! Banyak hadiahnya lho!… https://t.co/wo36HBVFi8","truncated":true,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/wo36HBVFi8","expanded\_url":"https://twitter.com/i/web/status/927385091889860609","display\_url":"twitter.com/i/web/status/9…","indices":[111,134]}]},"source":"<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">Hootsuite<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":1,"favorite\_count":2,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},**

**{"created\_at":"Mon Nov 06 03:21:32 +0000 2017","id":9.2737536854093e+017,"id\_str":"927375368540930050","text":"Saksikan pertandingannya mulai Desember 2017 di channel UseePrime no 88 #myIndiHomeIBL #RaiseUP #IBL https://t.co/sjqeMdTSCK","truncated":false,"entities":{"hashtags":[{"text":"myIndiHomeIBL","indices":[72,86]},{"text":"RaiseUP","indices":[87,95]},{"text":"IBL","indices":[96,100]}],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/sjqeMdTSCK","expanded\_url":"https://twitter.com/TelkomIndonesia/status/927370883492077568","display\_url":"twitter.com/TelkomIndonesi…","indices":[101,124]}]},"source":"<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":true,"quoted\_status\_id":9.27370883492078e+017,"quoted\_status\_id\_str":"927370883492077568","quoted\_status":{"created\_at":"Mon Nov 06 03:03:42 +0000 2017","id":9.27370883492078e+017,"id\_str":"927370883492077568","text":"Telkom Indonesia menjadi sponsor pd kompetisi basket terbesar di Indonesia. Selengkapnya cek di… https://t.co/A6mlrLdmmk","truncated":true,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/A6mlrLdmmk","expanded\_url":"https://twitter.com/i/web/status/927370883492077568","display\_url":"twitter.com/i/web/status/9…","indices":[97,120]}]},"source":"<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":80204488,"id\_str":"80204488"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":7,"favorite\_count":12,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},"retweet\_count":1,"favorite\_count":1,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},**

**{"created\_at":"Mon Nov 06 02:33:30 +0000 2017","id":9.27363280749949e+017,"id\_str":"927363280749948928","text":"Aplikasi myIndiHome banyak benefitnya lho! Mau tau apa saja? Download skarang di Google Play&amp;App Store!… https://t.co/59RrG4SxrD","truncated":true,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/59RrG4SxrD","expanded\_url":"https://twitter.com/i/web/status/927363280749948928","display\_url":"twitter.com/i/web/status/9…","indices":[109,132]}]},"source":"<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">Hootsuite<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":1,"favorite\_count":2,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"},**

**{"created\_at":"Sun Nov 05 12:00:32 +0000 2017","id":9.2714359239261e+017,"id\_str":"927143592392609792","text":"Ragam acara Korea dan Jepang kini semakin seru dengan hadirnya channel S One dan S Gem di UseeTV! Yuk saksikan dram… https://t.co/DVhu0igSxW","truncated":true,"entities":{"hashtags":[],"symbols":[],"user\_mentions":[],"urls":[{"url":"https://t.co/DVhu0igSxW","expanded\_url":"https://twitter.com/i/web/status/927143592392609792","display\_url":"twitter.com/i/web/status/9…","indices":[117,140]}]},"source":"<a href=\"http://www.hootsuite.com\" rel=\"nofollow\">Hootsuite<\/a>","in\_reply\_to\_status\_id":null,"in\_reply\_to\_status\_id\_str":null,"in\_reply\_to\_user\_id":null,"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_screen\_name":null,"user":{"id":528328843,"id\_str":"528328843"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"retweet\_count":3,"favorite\_count":1,"favorited":false,"retweeted":false,"possibly\_sensitive":false,"lang":"in"}**

**],**

**}**

**bungkus\_bracket = []**

**for element\_yeuh in datanyah['base']:**

**# if element\_yeuh['serial'] == '00000000762c1d3c':**

**# print 'there'**

**# print element\_yeuh['text']**

**bungkus\_bracket = element\_yeuh['text']**

**# print bungkus\_bracket**

**text = [element\_yeuh['text']]**

**kamu\_d = {}**

**for words\_ajah in datanyah['base']:**

**if words\_ajah:**

**if words\_ajah['text'][0] in kamu\_d: # Cek Apa ada karakter pertama di kamus yang cacahin**

**kamu\_d[words\_ajah['text'][0]] = kamu\_d[words\_ajah['text'][0]] + 1 # gunakan karakter pertama sebagai kunci ke kamus**

**else: # jika engga ada mari menghihtung**

**kamu\_d[words\_ajah['text'][0]] = 1 # gunakan karakter pertama sebagai kunci ke kamus**

**word\_counter\_eh = {}**

**for word\_ea in datanyah['base']:**

**if word\_ea['text'][0] in word\_counter\_eh:**

**word\_counter\_eh[word\_ea['text'][0]] += 1**

**else:**

**word\_counter\_eh[word\_ea['text'][0]] = 1**

**words\_terkenal = sorted(word\_counter\_eh, key = word\_counter\_eh.get, reverse = True)**

**top\_3\_terkenal = words\_terkenal[:3]**

**# baru sampe sini**

**print ("Top 3 Words: ", (top\_3\_terkenal))**

**print ("Word\_count: ", (kamu\_d))**

**# next -> hashtag\_count**

**# next -> mention\_count**

**# next -> top3\_hastagh**

**# next -> top3\_mention**